



**Europäisches
Patentamt**

**European
Patent Office**

**Office européen
des brevets**

CT/EP03/09145

REC'D 10 NOV 2003

WIPO

PCT

Bescheinigung

Certificate

Attestation

Die angehefteten Unterla-
gen stimmen mit der
ursprünglich eingereichten
Fassung der auf dem näch-
sten Blatt bezeichneten
europäischen Patentanmel-
dung überein.

The attached documents
are exact copies of the
European patent application
described on the following
page, as originally filed.

Les documents fixés à
cette attestation sont
conformes à la version
initialement déposée de
la demande de brevet
européen spécifiée à la
page suivante.

Patentanmeldung Nr. Patent application No. Demande de brevet n°

02102173.8

PRIORITY DOCUMENT
SUBMITTED OR TRANSMITTED IN
COMPLIANCE WITH
RULE 17.1(a) OR (b)

Der Präsident des Europäischen Patentamts;
Im Auftrag

For the President of the European Patent Office

Le Président de l'Office européen des brevets
p.o.

R C van Dijk



Europäisches
Patentamt

European
Patent Office

Office européen
des brevets

PCT/EP03 / 09145

Anmeldung Nr:

Application no.: 02102173.8

Demande no:

Anmeldetag:

Date of filing: 19.08.02

Date de dépôt:

Anmelder/Applicant(s)/Demandeur(s):

DSM N.V.

Het Overloon 1

6411 TE Heerlen

PAYS-BAS

Bezeichnung der Erfindung/Title of the invention/Titre de l'invention:

(Falls die Bezeichnung der Erfindung nicht angegeben ist, siehe Beschreibung.

If no title is shown please refer to the description.

Si aucun titre n'est indiqué se referer à la description.)

NOVEL BAKING ENZYME NBE 033 AND USES THEREOF

In Anspruch genommene Priorität(en) / Priority(ies) claimed /Priorité(s)
revendiquée(s)

Staat/Tag/Aktenzeichen/State/Date/File no./Pays/Date/Numéro de dépôt:

Internationale Patentklassifikation/International Patent Classification/
Classification internationale des brevets:

C12N9/00

Am Anmeldetag benannte Vertragstaaten/Contracting states designated at date of
filing/Etats contractants désignées lors du dépôt:

AT BE BG CH CY CZ DE DK EE ES FI FR GB GR IE IT LI LU MC NL PT SE SK TR

NOVEL BAKING ENZYME NBE 033 AND USES THEREOF

5

Field of the invention

The invention relates to a newly identified polynucleotide sequence comprising a gene that encodes a novel baking enzyme. The invention features the full length nucleotide sequence of the novel gene, the cDNA sequence comprising the full length coding sequence of the novel baking enzyme as well as the amino acid sequence of the full-length functional protein and functional equivalents thereof. The invention also relates to methods of using these enzymes in industrial processes and methods of diagnosing fungal infections. Also included in the invention are cells transformed with a polynucleotide according to the invention and cells wherein a baking enzyme according to the invention is genetically modified to enhance its activity and/or level of expression.

Background of the invention

20

Baked products such as bread are prepared from a dough which is usually made from the basic ingredients (wheat) flour, water and optionally salt. Depending on the baked products, other ingredients added may be are sugars, flavours etceteras. For leavened products, primarily baker's yeast is used next to chemical leavening systems such as a combination of an acid (generating compound) and bicarbonate.

25

In order to improve the handling properties of the dough and/or the final properties of the baked products there is a continuous effort to develop processing aids with improving properties. Processing aids are defined herein as compounds that improve the handling properties of the dough and/or the final properties of the baked products. Dough properties that may be improved comprise machineability, gas retaining capability, reduced stickiness, elasticity, extensibility, moldability etcetera. Properties of the baked products that may be improved comprise loaf volume, crust crispiness, crumb texture and softness, flavour relative staleness and shelf life. These dough and/or baked product improving processing aids can be divided into two groups: chemical additives and enzymes.

30
35

Yeast, enzymes and chemical additives are generally added separately to the dough. Yeast may be added as a liquid suspension, in a compressed form or as

active dry (ADY) or instant dry yeast (IDY). The difference between these yeast formulations is the water- and yeast dry matter content. Liquid yeast has a yeast dry matter content of less than 25% (w/v). Cream yeast is a particular form of liquid yeast and has a dry matter content between 17 and 23% (w/v). Compressed yeast has a dry matter content between 25-35% (w/v) while the dry yeast formulations have a dry matter content between 92-98% (w/v).

Enzymes may be added in a dry, e.g. granulated form or in dissolved form. The chemical additives are in most cases added in powder form. Also, processing aid compositions which are tailored to specific baking applications, may be composed of a dedicated mixture of chemical additives and enzyme.

The preparation of a dough from the ingredients and processing aids described above is well known in the art and comprises mixing of said ingredients and processing aids and one or more moulding and fermentation steps.

The preparation of baked products from such doughs is also well known in the art and may comprise molding and shaping and further fermentation of the dough followed by baking at required temperatures and baking times.

Chemical additives with improving properties comprise oxidising agents such as ascorbic acid, bromate and azodicarbonate, reducing agents such as L-cysteine and glutathione, emulsifiers acting as dough conditioners such as diacetyl tartaric esters of mono/diglycerides (DATEM), sodium stearyl lactylate (SSL) or calcium stearyl lactylate (CSL), or acting as crumb softeners such as glycerol monostearate (GMS) etceteras, fatty materials such as triglycerides (fat) or lecithin and others.

As a result of a consumer-driven need to replace the chemical additives by more natural products, several enzymes have been developed with dough and/or baked product improving properties and which are used in all possible combinations depending on the specific baking application conditions. Suitable enzymes include starch degrading enzymes, arabinoxylan- and other hemicellulose degrading enzymes, cellulose degrading enzymes, oxidizing enzymes, fatty material splitting enzymes, protein degrading, modifying or crosslinking enzymes.

Starch degrading enzymes are for instance endo-acting enzymes such as alpha-amylase, maltogenic amylase, pullulanase or other debranching enzymes and exo-acting enzymes that cleave off glucose (amyloglucosidase), maltose (beta-amylase), maltotriose, maltotetraose and higher oligosaccharides.

Arabinoxylan- and other hemicellulose degrading enzymes are for instance xylanases, pentosanases, hemicellulase, arabinofuranosidase, glucanase and

others.

Cellulose degrading enzymes are for instance cellulase (beta-1,4-endoglucanase), cellobiohydrolase and beta-glucosidase.

5 Oxidizing enzymes are for instance glucose oxidase, hexose oxidase, pyranose oxidase, sulfhydryl oxidase, lipoxygenase, laccase, polyphenol oxidases and others.

Fatty material splitting enzymes are for instance triacylglycerol lipases, phospholipases (such as A₁, A₂, B, C and D) and galactolipases.

10 Protein degrading, modifying or crosslinking enzymes are for instance endo-acting proteases (serine proteases, metalloproteases, aspartyl proteases, thiol proteases), exo-acting peptidases that cleave off one amino acid, or dipeptide, tripeptide etceteras from the N-terminal (aminopeptidases) or C-terminal (carboxypeptidases) ends of the polypeptide chain, asparagines or glutamine deamidating enzymes such as deamidase and peptidoglutaminase or crosslinking enzymes such as transglutaminase.

15 Baking enzymes may conveniently be produced in microorganisms. Microbial baking enzymes are available from a variety of sources; Bacillus spec. are a common source of bacterial enzymes, whereas fungal enzymes are commonly produced in Aspergillus spec.

20 Baking enzymes may be used in a manifold of baked goods. The term "baked goods" is herein defined as to comprise bread products such as tin bread, loaves of bread, French bread as well as rolls, cakes, pies, muffins, yeast raised and cake doughnuts and the like.

25 In the above processes, it is advantageous to use baking enzymes that are obtained by recombinant DNA techniques. Such recombinant enzymes have a number of advantages over their traditionally purified counterparts. Recombinant enzymes may be produced at a low cost price, high yield, free from contaminating agents like bacteria or viruses but also free from bacterial toxins or contaminating other enzyme activities.

Object of the invention

30

It is an object of the invention to provide novel polynucleotides encoding novel baking enzymes with improved properties. A further object is to provide naturally and recombinantly produced baking enzymes as well as recombinant strains producing these. Also fusion polypeptides are part of the invention as well as
35 methods of making and using the polynucleotides and polypeptides according to the invention.

Summary of the invention

The invention provides for novel polynucleotides encoding novel baking enzymes, in particular enzymes with Triacylglycerol lipase activity (E.C.

5 Number 3.1.1.3).

More in particular, the invention provides for polynucleotides having a nucleotide sequence that hybridises preferably under highly stringent conditions to a sequence according to SEQ ID NO: 1 or SEQ ID NO: 2. Consequently, the invention provides nucleic acids that are more than 40% such as about 60%, preferably 65%,
10 more preferably 70%, even more preferably 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% homologous to the sequences according to SEQ ID NO: 1 or SEQ ID NO: 2.

In a more preferred embodiment the invention provides for such an isolated polynucleotide obtainable from a filamentous fungus, in particular *A. niger* is
15 preferred.

In one embodiment, the invention provides for an isolated polynucleotide comprising a nucleic acid sequence encoding a polypeptide with an amino acid sequence as shown in SEQ ID NO: 3 or functional equivalents thereof.

In a further preferred embodiment, the invention provides an isolated
20 polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

In a preferred embodiment the invention provides a baking enzyme gene according to SEQ ID NO: 1. In another aspect the invention provides a polynucleotide, preferably a cDNA encoding an *A. niger* baking enzyme whose amino
25 acid sequence is shown in SEQ ID NO: 3 or variants or fragments of that polypeptide. In a preferred embodiment the cDNA has a sequence according to SEQ ID NO: 2 or functional equivalents thereof.

In an even further preferred embodiment, the invention provides for a polynucleotide comprising the coding sequence of the polynucleotides according to the
30 invention, preferred is the polynucleotide sequence of SEQ ID NO: 2.

The invention also relates to vectors comprising a polynucleotide sequence according to the invention and primers, probes and fragments that may be used to amplify or detect the DNA according to the invention.

In a further preferred embodiment, a vector is provided wherein the
35 polynucleotide sequence according to the invention is functionally linked with regulatory sequences suitable for expression of the encoded amino acid sequence in a suitable

host cell, such as *A. niger* or *A. oryzae*. The invention also provides methods for preparing polynucleotides and vectors according to the invention.

The invention also relates to recombinantly produced host cells that contain heterologous or homologous polynucleotides according to the invention.

5 In another embodiment, the invention provides recombinant host cells wherein the expression of a baking enzyme according to the invention is significantly increased or wherein the activity of the baking enzyme is increased.

10 In another embodiment the invention provides for a recombinantly produced host cell that contains heterologous or homologous DNA according to the invention and wherein the cell is capable of producing a functional baking enzyme according to the invention, preferably a cell capable of over-expressing the baking enzyme according to the invention, for example an *Aspergillus* strain comprising an increased copy number of a gene or cDNA according to the invention.

15 In yet another aspect of the invention, a purified polypeptide is provided. The polypeptides according to the invention include the polypeptides encoded by the polynucleotides according to the invention. Especially preferred is a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

20 Accordingly, in one aspect the present invention provides a baking enzyme composition containing as an active ingredient an enzyme according to SEQ ID NO: 3 or functional equivalents thereof.

In another aspect, the invention provides a method of making baked goods wherein there is incorporated into the dough used for making the baked goods an enzyme according to SEQ ID NO: 3 or functional equivalents thereof.

25 Fusion proteins comprising a polypeptide according to the invention are also within the scope of the invention. The invention also provides methods of making the polypeptides according to the invention.

The invention also relates to the use of the baking enzyme according to the invention in any industrial process as described herein

30 **Detailed description of the invention**

Polynucleotides

35 The present invention provides polynucleotides encoding a baking enzyme, tentatively called NBE 033, having an amino acid sequence according to SEQ ID NO: 3 or functional equivalents thereof. The sequence of the gene encoding NBE

033 was determined by sequencing a genomic clone obtained from *Aspergillus niger*. The invention provides polynucleotide sequences comprising the gene encoding the NBE 033 baking enzyme as well as its complete cDNA sequence and its coding sequence. Accordingly, the invention relates to an isolated polynucleotide comprising the nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2 or functional equivalents thereof.

More in particular, the invention relates to an isolated polynucleotide hybridisable under stringent conditions, preferably under highly stringent conditions, to a polynucleotide according to SEQ ID NO: 1 or SEQ ID NO: 2. Advantageously, such polynucleotides may be obtained from filamentous fungi, in particular from *Aspergillus niger*. More specifically, the invention relates to an isolated polynucleotide having a nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2.

The invention also relates to an isolated polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.

As used herein, the terms "gene" and "recombinant gene" refer to nucleic acid molecules which may be isolated from chromosomal DNA, which include an open reading frame encoding a protein, e.g. an *A. niger* baking enzyme. A gene may include coding sequences, non-coding sequences, introns and regulatory sequences. Moreover, a gene refers to an isolated nucleic acid molecule as defined herein.

A nucleic acid molecule of the present invention, such as a nucleic acid molecule having the nucleotide sequence of SEQ ID NO: 1 or SEQ ID NO: 2 or a functional equivalent thereof, can be isolated using standard molecular biology techniques and the sequence information provided herein. For example, using all or portion of the nucleic acid sequence of SEQ ID NO: 1 or the nucleotide sequence of SEQ ID NO: 2 as a hybridization probe, nucleic acid molecules according to the invention can be isolated using standard hybridization and cloning techniques (e. g., as described in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual*. 2nd, ed., Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Moreover, a nucleic acid molecule encompassing all or a portion of SEQ ID NO: 1 or SEQ ID NO: 2 can be isolated by the polymerase chain reaction (PCR) using synthetic oligonucleotide primers designed based upon the sequence information contained in SEQ ID NO:1 or SEQ ID NO: 2.

A nucleic acid of the invention can be amplified using cDNA, mRNA

or alternatively, genomic DNA, as a template and appropriate oligonucleotide primers according to standard PCR amplification techniques. The nucleic acid so amplified can be cloned into an appropriate vector and characterized by DNA sequence analysis.

Furthermore, oligonucleotides corresponding to or hybridisable to
5 nucleotide sequences according to the invention can be prepared by standard synthetic techniques, e. g., using an automated DNA synthesizer.

In a preferred embodiment, an isolated nucleic acid molecule of the invention comprises the nucleotide sequence shown in SEQ ID NO: 2. The sequence of SEQ ID NO: 2 corresponds to the coding region of the A. niger NBE 033 cDNA. This
10 cDNA comprises sequences encoding the A. niger NBE 033 polypeptide according to SEQ ID NO: 3.

In another preferred embodiment, an isolated nucleic acid molecule of the invention comprises a nucleic acid molecule which is a complement of the nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO: 2 or a functional
15 equivalent of these nucleotide sequences.

A nucleic acid molecule which is complementary to another nucleotide sequence is one which is sufficiently complementary to the other nucleotide sequence such that it can hybridize to the other nucleotide sequence thereby forming a stable duplex.

20 One aspect of the invention pertains to isolated nucleic acid molecules that encode a polypeptide of the invention or a functional equivalent thereof such as a biologically active fragment or domain, as well as nucleic acid molecules sufficient for use as hybridisation probes to identify nucleic acid molecules encoding a polypeptide of the invention and fragments of such nucleic acid molecules suitable for
25 use as PCR primers for the amplification or mutation of nucleic acid molecules.

An "isolated polynucleotide" or "isolated nucleic acid" is a DNA or RNA that is not immediately contiguous with both of the coding sequences with which it is immediately contiguous (one on the 5' end and one on the 3' end) in the naturally occurring genome of the organism from which it is derived. Thus, in one embodiment,
30 an isolated nucleic acid includes some or all of the 5' non-coding (e.g., promotor) sequences that are immediately contiguous to the coding sequence. The term therefore includes, for example, a recombinant DNA that is incorporated into a vector, into an autonomously replicating plasmid or virus, or into the genomic DNA of a prokaryote or eukaryote, or which exists as a separate molecule (e.g., a cDNA or a
35 genomic DNA fragment produced by PCR or restriction endonuclease treatment) independent of other sequences. It also includes a recombinant DNA that is part of a

hybrid gene encoding an additional polypeptide that is substantially free of cellular material, viral material, or culture medium (when produced by recombinant DNA techniques), or chemical precursors or other chemicals (when chemically synthesized). Moreover, an "isolated nucleic acid fragment" is a nucleic acid fragment that is not naturally occurring as a fragment and would not be found in the natural state.

As used herein, the terms "polynucleotide" or "nucleic acid molecule" are intended to include DNA molecules (e.g., cDNA or genomic DNA) and RNA molecules (e.g., mRNA) and analogs of the DNA or RNA generated using nucleotide analogs. The nucleic acid molecule can be single-stranded or double-stranded, but preferably is double-stranded DNA. The nucleic acid may be synthesized using oligonucleotide analogs or derivatives (e.g., inosine or phosphorothioate nucleotides). Such oligonucleotides can be used, for example, to prepare nucleic acids that have altered base-pairing abilities or increased resistance to nucleases.

Another embodiment of the invention provides an isolated nucleic acid molecule which is antisense to an NBE 033 nucleic acid molecule, e.g., the coding strand of an NBE 033 nucleic acid molecule. Also included within the scope of the invention are the complement strands of the nucleic acid molecules described herein.

Sequencing errors

The sequence information as provided herein should not be so narrowly construed as to require inclusion of erroneously identified bases. The specific sequences disclosed herein can be readily used to isolate the complete gene from filamentous fungi, in particular *A. niger* which in turn can easily be subjected to further sequence analyses thereby identifying sequencing errors.

Unless otherwise indicated, all nucleotide sequences determined by sequencing a DNA molecule herein were determined using an automated DNA sequencer and all amino acid sequences of polypeptides encoded by DNA molecules determined herein were predicted by translation of a DNA sequence determined as above. Therefore, as is known in the art for any DNA sequence determined by this automated approach, any nucleotide sequence determined herein may contain some errors. Nucleotide sequences determined by automation are typically at least about 90% identical, more typically at least about 95% to at least about 99.9% identical to the actual nucleotide sequence of the sequenced DNA molecule. The actual sequence can be more precisely determined by other approaches including manual DNA sequencing methods well known in the art. As is also known in the art, a single insertion or deletion

in a determined nucleotide sequence compared to the actual sequence will cause a frame shift in translation of the nucleotide sequence such that the predicted amino acid sequence encoded by a determined nucleotide sequence will be completely different from the amino acid sequence actually encoded by the sequenced DNA molecule, beginning at the point of such an insertion or deletion.

The person skilled in the art is capable of identifying such erroneously identified bases and knows how to correct for such errors.

Nucleic acid fragments, probes and primers

A nucleic acid molecule according to the invention may comprise only a portion or a fragment of the nucleic acid sequence shown in SEQ ID NO:1 or SEQ ID NO:2, for example a fragment which can be used as a probe or primer or a fragment encoding a portion of a NBE 033 protein. The nucleotide sequence determined from the cloning of the NBE 033 gene and cDNA allows for the generation of probes and primers designed for use in identifying and/or cloning other NBE 033 family members, as well as NBE 033 homologues from other species. The probe/primer typically comprises substantially purified oligonucleotide which typically comprises a region of nucleotide sequence that hybridizes preferably under highly stringent conditions to at least about 12 or 15, preferably about 18 or 20, preferably about 22 or 25, more preferably about 30, 35, 40, 45, 50, 55, 60, 65, or 75 or more consecutive nucleotides of a nucleotide sequence shown in SEQ ID NO:1 or SEQ ID NO: 2 or of a functional equivalent thereof.

Probes based on the NBE 033 nucleotide sequences can be used to detect transcripts or genomic NBE 033 sequences encoding the same or homologous proteins for instance in other organisms. In preferred embodiments, the probe further comprises a label group attached thereto, e.g., the label group can be a radioisotope, a fluorescent compound, an enzyme, or an enzyme cofactor. Such probes can also be used as part of a diagnostic test kit for identifying cells which express a NBE 033 protein.

Identity & homology

The terms "homology" or "percent identity" are used interchangeably herein. For the purpose of this invention, it is defined here that in order to determine the percent identity of two amino acid sequences or of two nucleic acid sequences, the

sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in the sequence of a first amino acid or nucleic acid sequence for optimal alignment with a second amino or nucleic acid sequence). The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position. The percent identity between the two sequences is a function of the number of identical positions shared by the sequences (i.e., % identity = number of identical positions/total number of positions (i.e. overlapping positions) x 100). Preferably, the two sequences are the same length.

The skilled person will be aware of the fact that several different computer programs are available to determine the homology between two sequences. For instance, a comparison of sequences and determination of percent identity between two sequences can be accomplished using a mathematical algorithm. In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (J. Mol. Biol. (48):444-453 (1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at <http://www.gcg.com>), using either a Blossum 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. The skilled person will appreciate that all these different parameters will yield slightly different results but that the overall percentage identity of two sequences is not significantly altered when using different algorithms.

In yet another embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (available at <http://www.gcg.com>), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity two amino acid or nucleotide sequence is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0) (available at: <http://vega.igh.cnrs.fr/bin/align-guess.cgi>) using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against public databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (1990) J. Mol. Biol. 215:403—10. BLAST nucleotide searches can be performed with

the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to NBE 033 nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to NBE 033 protein molecules of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al., (1997) Nucleic Acids Res. 25(17):3389-3402. When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used. See <http://www.ncbi.nlm.nih.gov>.

Hybridisation

As used herein, the term "hybridizing" is intended to describe conditions for hybridization and washing under which nucleotide sequences at least about 50%, at least about 60%, at least about 70%, more preferably at least about 80%, even more preferably at least about 85% to 90%, more preferably at least 95% homologous to each other typically remain hybridized to each other.

A preferred, non-limiting example of such hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about 45 °C, followed by one or more washes in 1 X SSC, 0.1 % SDS at 50 °C, preferably at 55 °C, preferably at 60 °C and even more preferably at 65 °C.

Highly stringent conditions include, for example, hybridizing at 68 °C in 5x SSC/5x Denhardt's solution / 1.0% SDS and washing in 0.2x SSC/0.1% SDS at room temperature. Alternatively, washing may be performed at 42 °C.

The skilled artisan will know which conditions to apply for stringent and highly stringent hybridisation conditions. Additional guidance regarding such conditions is readily available in the art, for example, in Sambrook et al., 1989, Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, N.Y.; and Ausubel et al. (eds.), 1995, Current Protocols in Molecular Biology, (John Wiley & Sons, N.Y.).

Of course, a polynucleotide which hybridizes only to a poly A sequence (such as the 3' terminal poly(A) tract of mRNAs), or to a complementary stretch of T (or U) residues, would not be included in a polynucleotide of the invention used to specifically hybridize to a portion of a nucleic acid of the invention, since such a polynucleotide would hybridize to any nucleic acid molecule containing a poly (A) stretch or the complement thereof (e.g., practically any double-standed cDNA clone).

Obtaining full length DNA from other organisms

5 In a typical approach, cDNA libraries constructed from other organisms, e.g. filamentous fungi, in particular from the species *Aspergillus* can be screened.

For example, *Aspergillus* strains can be screened for homologous NBE 033 polynucleotides by Northern blot analysis. Upon detection of transcripts homologous to polynucleotides according to the invention, cDNA libraries can be
10 constructed from RNA isolated from the appropriate strain, utilizing standard techniques well known to those of skill in the art. Alternatively, a total genomic DNA library can be screened using a probe hybridisable to a NBE 033 polynucleotide according to the invention.

Homologous gene sequences can be isolated, for example, by
15 performing PCR using two degenerate oligonucleotide primer pools designed on the basis of nucleotide sequences as taught herein.

The template for the reaction can be cDNA obtained by reverse transcription of mRNA prepared from strains known or suspected to express a polynucleotide according to the invention. The PCR product can be subcloned and
20 sequenced to ensure that the amplified sequences represent the sequences of a new NBE 033 nucleic acid sequence, or a functional equivalent thereof.

The PCR fragment can then be used to isolate a full length cDNA clone by a variety of known methods. For example, the amplified fragment can be labeled and used to screen a bacteriophage or cosmid cDNA library. Alternatively, the
25 labeled fragment can be used to screen a genomic library.

PCR technology also can be used to isolate full length cDNA sequences from other organisms. For example, RNA can be isolated, following standard procedures, from an appropriate cellular or tissue source. A reverse transcription reaction can be performed on the RNA using an oligonucleotide primer
30 specific for the most 5' end of the amplified fragment for the priming of first strand synthesis.

The resulting RNA/DNA hybrid can then be "tailed" (e.g., with guanines) using a standard terminal transferase reaction, the hybrid can be digested with RNase H, and second strand synthesis can then be primed (e.g., with a poly-C
35 primer). Thus, cDNA sequences upstream of the amplified fragment can easily be isolated. For a review of useful cloning strategies, see e.g., Sambrook et al., supra; and

Ausubel et al., supra.

Whether or not a homologous DNA fragment encodes a functional NBE 033 protein, may easily be tested by methods known in the art. Methods for testing enzyme activity of Triacylglycerol lipase s (E.C. number 3.1.1.3) may be found
5 on <http://www.biochem.ucl.ac.uk/bsm/enzymes/>.

Vectors

Another aspect of the invention pertains to vectors, preferably
10 expression vectors, containing a nucleic acid encoding a NBE 033 protein or a functional equivalent thereof. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector,
15 wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (e.g., bacterial vectors having a bacterial origin of replication and episomal mammalian vectors). Other vectors (e.g., non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated
20 along with the host genome. Moreover, certain vectors are capable of directing the expression of genes to which they are operatively linked. Such vectors are referred to herein as "expression vectors". In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. The terms "plasmid" and "vector" can be used interchangeably herein as the plasmid is the most commonly used form of
25 vector. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (e.g., replication defective retroviruses, adenoviruses and adeno-associated viruses), which serve equivalent functions.

The recombinant expression vectors of the invention comprise a nucleic acid of the invention in a form suitable for expression of the nucleic acid in a
30 host cell, which means that the recombinant expression vector includes one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operatively linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner
35 which allows for expression of the nucleotide sequence (e.g., in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the

host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (e.g., polyadenylation signal). Such regulatory sequences are described, for example, in Goeddel; *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cells and those which direct expression of the nucleotide sequence only in a certain host cell (e.g. tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed, the level of expression of protein desired, etc. The expression vectors of the invention can be introduced into host cells to thereby produce proteins or peptides, encoded by nucleic acids as described herein (e.g. NBE 033 proteins, mutant forms of NBE 033 proteins, fragments, variants or functional equivalents thereof, fusion proteins, etc.).

The recombinant expression vectors of the invention can be designed for expression of NBE 033 proteins in prokaryotic or eukaryotic cells. For example, NBE 033 proteins can be expressed in bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors) yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Expression vectors useful in the present invention include chromosomal-, episomal- and virus-derived vectors e.g., vectors derived from bacterial plasmids, bacteriophage, yeast episome, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations thereof, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids.

The DNA insert should be operatively linked to an appropriate promoter, such as the phage lambda PL promoter, the *E. coli* lac, trp and tac promoters, the SV40 early and late promoters and promoters of retroviral LTRs, to name a few. Other suitable promoters will be known to the skilled person. In a specific embodiment, promoters are preferred that are capable of directing a high expression level of baking enzymes in filamentous fungi. Such promoters are known in the art. The expression constructs may contain sites for transcription initiation, termination, and, in the transcribed region, a ribosome binding site for translation. The coding portion of the

mature transcripts expressed by the constructs will include a translation initiating AUG at the beginning and a termination codon appropriately positioned at the end of the polypeptide to be translated.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing foreign nucleic acid (e.g., DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, transduction, infection, lipofection, cationic lipid-mediated transfection or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989), Davis et al., *Basic Methods in Molecular Biology* (1986) and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (e.g., resistance to antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those which confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid encoding a selectable marker can be introduced into a host cell on the same vector as that encoding a NBE 033 protein or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid can be identified by drug selection (e.g. cells that have incorporated the selectable marker gene will survive, while the other cells die).

Expression of proteins in prokaryotes is often carried out in *E. coli* with vectors containing constitutive or inducible promoters directing the expression of either fusion or non-fusion proteins. Fusion vectors add a number of amino acids to a protein encoded therein, e.g. to the amino terminus of the recombinant protein. Such fusion vectors typically serve three purposes: 1) to increase expression of recombinant protein; 2) to increase the solubility of the recombinant protein; and 3) to aid in the purification of the recombinant protein by acting as a ligand in affinity purification. Often, in fusion expression vectors, a proteolytic cleavage site is introduced at the junction of the fusion moiety and the recombinant protein to enable separation of the recombinant protein from the fusion moiety subsequent to purification of the fusion protein. Such enzymes, and their cognate recognition sequences, include Factor Xa,

thrombin and enterokinase.

As indicated, the expression vectors will preferably contain selectable markers. Such markers include dihydrofolate reductase or neomycin resistance for eukarotic cell culture and tetracycline or ampicilling resistance for culturing in *E. coli* and other bacteria. Representative examples of appropriate host include bacterial cells, such as *E. coli*, *Streptomyces* and *Salmonella typhimurium*; fungal cells, such as yeast; insect cells such as *Drosophila* S2 and *Spodoptera* Sf9; animal cells such as CHO, COS and Bowes melanoma; and plant cells. Appropriate culture mediums and conditions for the above-described host cells are known in the art.

Among vectors preferred for use in bacteria are pQE70, pQE60 and pQE-9; available from Qiagen; pBS vectors, Phagescript vectors, Bluescript vectors, pNH8A, pNH16A, pNH18A, pNH46A, available from Stratagene; and ptrc99a, pKK223-3, pKK233-3, pDR540, pRIT5 available from Pharmacia. Among preferred eukaryotic vectors are PWLNEO, pSV2CAT, pOG44, pZT1 and pSG available from Stratagene; and pSVK3, pBPV, pMSG and pSVL available from Pharmacia. Other suitable vectors will be readily apparent to the skilled artisan.

Among known bacterial promoters for use in the present invention include *E. coli* lacI and lacZ promoters, the T3 and T7 promoters, the gpt promoter, the lambda PR, PL promoters and the trp promoter, the HSV thymidine kinase promoter, the early and late SV40 promoters, the promoters of retroviral LTRs, such as those of the Rous sarcoma virus ("RSV"), and metallothionein promoters, such as the mouse metallothionein-I promoter.

Transcription of the DNA encoding the polypeptides of the present invention by higher eukaryotes may be increased by inserting an enhancer sequence into the vector. Enhancers are cis-acting elements of DNA, usually about from 10 to 300 bp that act to increase transcriptional activity of a promoter in a given host cell-type. Examples of enhancers include the SV40 enhancer, which is located on the late side of the replication origin at bp 100 to 270, the cytomegalovirus early promoter enhancer, the polyoma enhancer on the late side of the replication origin, and adenovirus enhancers.

For secretion of the translated protein into the lumen of the endoplasmic reticulum, into the periplasmic space or into the extracellular environment, appropriate secretion signal may be incorporated into the expressed polypeptide. The signals may be endogenous to the polypeptide or they may be heterologous signals.

The polypeptide may be expressed in a modified form, such as a fusion protein, and may include not only secretion signals but also additional

heterologous functional regions. Thus, for instance, a region of additional amino acids, particularly charged amino acids, may be added to the N-terminus of the polypeptide to improve stability and persistence in the host cell, during purification or during subsequent handling and storage. Also, peptide moieties may be added to the polypeptide to facilitate purification.

Polypeptides according to the invention

The invention provides an isolated polypeptide having the amino acid sequence according to SEQ ID NO: 3, an amino acid sequence obtainable by expressing the polynucleotide of SEQ ID NO: 1 in an appropriate host, as well as an amino acid sequence obtainable by expressing the polynucleotide sequences of SEQ ID NO: 2 in an appropriate host. Also, a peptide or polypeptide comprising a functional equivalent of the above polypeptides is comprised within the present invention. The above polypeptides are collectively comprised in the term "polypeptides according to the invention"

The terms "peptide" and "oligopeptide" are considered synonymous (as is commonly recognized) and each term can be used interchangeably as the context requires to indicate a chain of at least two amino acids coupled by peptidyl linkages. The word "polypeptide" is used herein for chains containing more than seven amino acid residues. All oligopeptide and polypeptide formulas or sequences herein are written from left to right and in the direction from amino terminus to carboxy terminus. The one-letter code of amino acids used herein is commonly known in the art and can be found in Sambrook, et al. (*Molecular Cloning: A Laboratory Manual*, 2nd, ed. Cold Spring Harbor Laboratory, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989)

By "isolated" polypeptide or protein is intended a polypeptide or protein removed from its native environment. For example, recombinantly produced polypeptides and proteins expressed in host cells are considered isolated for the purpose of the invention as are native or recombinant polypeptides which have been substantially purified by any suitable technique such as, for example, the single-step purification method disclosed in Smith and Johnson, Gene 67:31-40 (1988).

The NBE 033 baking enzyme according to the invention can be recovered and purified from recombinant cell cultures by well-known methods including ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction

chromatography, affinity chromatography, hydroxylapatite chromatography and lectin chromatography. Most preferably, high performance liquid chromatography ("HPLC") is employed for purification.

5 Polypeptides of the present invention include naturally purified products, products of chemical synthetic procedures, and products produced by recombinant techniques from a prokaryotic or eukaryotic host, including, for example, bacterial, yeast, higher plant, insect and mammalian cells. Depending upon the host employed in a recombinant production procedure, the polypeptides of the present invention may be glycosylated or may be non-glycosylated. In addition, polypeptides of
10 the invention may also include an initial modified methionine residue, in some cases as a result of host-mediated processes.

A NBE 033 polypeptide according to the invention may be advantageously used in baking processes. The amount of enzyme to be added to the dough is determined empirically. It may depend on the quality of the flour used, the
15 degree of improvement which is required, the kind of bread or baked goods, the method of preparing the dough, the proportion of other ingredients etcetera.

The effect of the addition of the NBE 033 enzyme according to the invention manifests itself in the physical properties of dough and the quality of the baked goods. By the addition of the enzyme, the dough has improved elasticity, is less
20 sticky, and machinability is significantly improved. Also, the volume of the baked goods is increased, the crumb has a finer structure and a better softness is obtained. Staling of the bread is suppressed when a NBE 033 enzyme is added to the dough.

Protein fragments

25 The invention also features biologically active fragments of the polypeptides according to the invention.

Biologically active fragments of a polypeptide of the invention include polypeptides comprising amino acid sequences sufficiently identical to or derived from
30 the amino acid sequence of the NBE 033 protein (e.g., the amino acid sequence of SEQ ID NO: 3), which include fewer amino acids than the full length protein, and exhibit at least one biological activity of the corresponding full-length protein. Typically, biologically active fragments comprise a domain or motif with at least one activity of the NBE 033 protein. Preferred is a fragment with Triacylglycerol lipase activity.

35 A biologically active fragment of a protein of the invention can be a polypeptide which is, for example, 10, 25, 50, 100 or more amino acids in length.

Moreover, other biologically active portions, in which other regions of the protein are deleted, can be prepared by recombinant techniques and evaluated for one or more of the biological activities of the native form of a polypeptide of the invention.

5 The invention also features nucleic acid fragments which encode the above biologically active fragments of the NBE 033 protein.

Fusion proteins

10 The proteins of the present invention or functional equivalents thereof, e.g., biologically active portions thereof, can be operatively linked to a non-NBE 033 polypeptide (e.g., heterologous amino acid sequences) to form fusion proteins. As used herein, a NBE 033 "chimeric protein" or "fusion protein" comprises a NBE 033 polypeptide operatively linked to a non-NBE 033 polypeptide. A "NBE 033 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to
15 NBE 033, whereas a "non-NBE 033 polypeptide" refers to a polypeptide having an amino acid sequence corresponding to a protein which is not substantially homologous to the NBE 033 protein, e.g., a protein which is different from the NBE 033 protein and which is derived from the same or a different organism. Within a NBE 033 fusion protein the NBE 033 polypeptide can correspond to all or a portion of a NBE 033
20 protein. In a preferred embodiment, a NBE 033 fusion protein comprises at least one biologically active fragment of a NBE 033 protein. In another preferred embodiment, a NBE 033 fusion protein comprises at least two biologically active portions of a NBE 033 protein. Within the fusion protein, the term "operatively linked" is intended to indicate that the NBE 033 polypeptide and the non-NBE 033 polypeptide are fused in-frame to
25 each other. The non-NBE 033 polypeptide can be fused to the N-terminus or C-terminus of the NBE 033 polypeptide.

For example, in one embodiment, the fusion protein is a GST-NBE 033 fusion protein in which the NBE 033 sequences are fused to the C-terminus of the GST sequences. Such fusion proteins can facilitate the purification of recombinant NBE
30 033. In another embodiment, the fusion protein is a NBE 033 protein containing a heterologous signal sequence at its N-terminus. In certain host cells (e.g., mammalian and Yeast host cells), expression and/or secretion of NBE 033 can be increased through use of a heterologous signal sequence.

35 In another example, the gp67 secretory sequence of the baculovirus envelope protein can be used as a heterologous signal sequence (*Current Protocols in Molecular Biology*, Ausubel et al., eds., John Wiley & Sons, 1992). Other examples of

eukaryotic heterologous signal sequences include the secretory sequences of melittin and human placental alkaline phosphatase (Stratagene; La Jolla, California). In yet another example, useful prokaryotic heterologous signal sequences include the phoA secretory signal (Sambrook et al., *supra*) and the protein A secretory signal (Pharmacia Biotech; Piscataway, New Jersey).

5 A signal sequence can be used to facilitate secretion and isolation of a protein or polypeptide of the invention. Signal sequences are typically characterized by a core of hydrophobic amino acids which are generally cleaved from the mature protein during secretion in one or more cleavage events. Such signal peptides contain
10 processing sites that allow cleavage of the signal sequence from the mature proteins as they pass through the secretory pathway. The signal sequence directs secretion of the protein, such as from a eukaryotic host into which the expression vector is transformed, and the signal sequence is subsequently or concurrently cleaved. The protein can then be readily purified from the extracellular medium by art recognized
15 methods. Alternatively, the signal sequence can be linked to the protein of interest using a sequence which facilitates purification, such as with a GST domain. Thus, for instance, the sequence encoding the polypeptide may be fused to a marker sequence, such as a sequence encoding a peptide, which facilitates purification of the fused polypeptide. In certain preferred embodiments of this aspect of the invention, the
20 marker sequence is a hexa-histidine peptide, such as the tag provided in a pQE vector (Qiagen, Inc.), among others, many of which are commercially available. As described in *Gentz et al, Proc. Natl. Acad. Sci. USA 86:821-824 (1989)*, for instance, hexa-histidine provides for convenient purification of the fusion protein. The HA tag is another peptide useful for purification which corresponds to an epitope derived of influenza
25 hemagglutinin protein, which has been described by Wilson *et al., Cell 37:767 (1984)*, for instance.

Preferably, a NBE 033 chimeric or fusion protein of the invention is produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in
30 accordance with conventional techniques, for example by employing blunt-ended or stagger-ended termini for ligation, restriction enzyme digestion to provide for appropriate termini, filling-in of cohesive ends as appropriate, alkaline phosphatase treatment to avoid undesirable joining, and enzymatic ligation. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated
35 DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two

consecutive gene fragments which can subsequently be annealed and reamplified to generate a chimeric gene sequence (see, for example, *Current Protocols in Molecular Biology*, eds. Ausubel *et al.* John Wiley & Sons: 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g, a GST polypeptide). A NBE 033-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the NBE 033 protein.

Functional equivalents

The terms "functional equivalents" and "functional variants" are used interchangeably herein. Functional equivalents of NBE 033 DNA are isolated DNA fragments that encode a polypeptide that exhibits a particular function of the NBE 033 A. niger baking enzyme as defined herein. A functional equivalent of a NBE 033 polypeptide according to the invention is a polypeptide that exhibits at least one function of an A. niger baking enzyme as defined herein. Functional equivalents therefore also encompass biologically active fragments.

Functional protein or polypeptide equivalents may contain only conservative substitutions of one or more amino acids of SEQ ID NO: 3 or substitutions, insertions or deletions of non-essential amino acids. Accordingly, a non-essential amino acid is a residue that can be altered in SEQ ID NO: 3 without substantially altering the biological function. For example, amino acid residues that are conserved among the NBE 033 proteins of the present invention, are predicted to be particularly unamenable to alteration. Furthermore, amino acids conserved among the NBE 033 proteins according to the present invention and other baking enzymes are not likely to be amenable to alteration.

The term "conservative substitution" is intended to mean that a substitution in which the amino acid residue is replaced with an amino acid residue having a similar side chain. These families are known in the art and include amino acids with basic side chains (e.g.lysine, arginine and hystidine), acidic side chains (e.g. aspartic acid, glutamic acid), uncharged polar side chains (e.g., glycine, asparagines, glutamine, serine, threonine, tyrosine, cysteine), non-polar side chains (e.g., alanine, valine, leucine, isoleucine, proline, phenylalanine, methionine, tryptophan), beta-branched side chains (e.g., threonine, valine, isoleucine) and aromatic side chains (e.g., tyrosine, phenylalanine tryptophan, histidine).

Functional nucleic acid equivalents may typically contain silent mutations or mutations that do not alter the biological function of encoded polypeptide.

Accordingly, the invention provides nucleic acid molecules encoding NBE 033 proteins that contain changes in amino acid residues that are not essential for a particular biological activity. Such NBE 033 proteins differ in amino acid sequence from SEQ ID NO: 3 yet retain at least one biological activity. In one embodiment the isolated nucleic acid molecule comprises a nucleotide sequence encoding a protein, wherein the protein comprises a substantially homologous amino acid sequence of at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO: 3.

For example, guidance concerning how to make phenotypically silent amino acid substitutions is provided in Bowie, J.U. et al., Science 247:1306-1310 (1990) wherein the authors indicate that there are two main approaches for studying the tolerance of an amino acid sequence to change. The first method relies on the process of evolution, in which mutations are either accepted or rejected by natural selection. The second approach uses genetic engineering to introduce amino acid changes at specific positions of a cloned gene and selects or screens to identify sequences that maintain functionality. As the authors state, these studies have revealed that proteins are surprisingly tolerant of amino acid substitutions. The authors further indicate which changes are likely to be permissive at a certain position of the protein. For example, most buried amino acid residues require non-polar side chains, whereas few features of surface side chains are generally conserved. Other such phenotypically silent substitutions are described in Bowie et al, supra, and the references cited therein.

An isolated nucleic acid molecule encoding a NBE 033 protein homologous to the protein according to SEQ ID NO: 3 can be created by introducing one or more nucleotide substitutions, additions or deletions into the coding nucleotide sequences according to SEQ ID NO: 1 or SEQ ID NO: 2 such that one or more amino acid substitutions, deletions or insertions are introduced into the encoded protein. Such mutations may be introduced by standard techniques, such as site-directed mutagenesis and PCR-mediated mutagenesis.

The term "functional equivalents" also encompasses orthologues of the *A. niger* NBE 033 protein. Orthologues of the *A. niger* NBE 033 protein are proteins that can be isolated from other strains or species and possess a similar or identical biological activity. Such orthologues can readily be identified as comprising an amino acid sequence that is substantially homologous to SEQ ID NO: 3.

As defined herein, the term "substantially homologous" refers to a first amino acid or nucleotide sequence which contains a sufficient or minimum number

of identical or equivalent (e.g., with similar side chain) amino acids or nucleotides to a second amino acid or nucleotide sequence such that the first and the second amino acid or nucleotide sequences have a common domain. For example, amino acid or nucleotide sequences which contain a common domain having about 60%, preferably
5 65%, more preferably 70%, even more preferably 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98% or 99% identity or more are defined herein as sufficiently identical.

Also, nucleic acids encoding other NBE 033 family members, which thus have a nucleotide sequence that differs from SEQ ID NO: 1 or SEQ ID NO: 2, are within the scope of the invention. Moreover, nucleic acids encoding NBE 033 proteins
10 from different species which thus have a nucleotide sequence which differs from SEQ ID NO: 1 or SEQ ID NO: 2 are within the scope of the invention.

Nucleic acid molecules corresponding to variants (e.g. natural allelic variants) and homologues of the NBE 033 DNA of the invention can be isolated based on their homology to the NBE 033 nucleic acids disclosed herein using the cDNAs
15 disclosed herein or a suitable fragment thereof, as a hybridisation probe according to standard hybridisation techniques preferably under highly stringent hybridisation conditions.

In addition to naturally occurring allelic variants of the NBE 033 sequence, the skilled person will recognise that changes can be introduced by mutation
20 into the nucleotide sequences of SEQ ID NO: 1 or SEQ ID NO: 2 thereby leading to changes in the amino acid sequence of the NBE 033 protein without substantially altering the function of the NBE 033 protein.

In another aspect of the invention, improved NBE 033 proteins are provided. Improved NBE 033 proteins are proteins wherein at least one biological
25 activity is improved. Such proteins may be obtained by randomly introducing mutations along all or part of the NBE 033 coding sequence, such as by saturation mutagenesis, and the resulting mutants can be expressed recombinantly and screened for biological activity. For instance, the art provides for standard assays for measuring the enzymatic activity of baking enzymes and thus improved proteins may easily be selected.

30 In a preferred embodiment the NBE 033 protein has an amino acid sequence according to SEQ ID NO: 3. In another embodiment, the NBE 033 polypeptide is substantially homologous to the amino acid sequence according to SEQ ID NO: 3 and retains at least one biological activity of a polypeptide according to SEQ ID NO: 3, yet differs in amino acid sequence due to natural variation or mutagenesis as
35 described above.

In a further preferred embodiment, the NBE 033 protein has an amino

acid sequence encoded by an isolated nucleic acid fragment capable of hybridising to a nucleic acid according to SEQ ID NO: 1 or SEQ ID NO: 2, preferably under highly stringent hybridisation conditions.

Accordingly, the NBE 033 protein is a protein which comprises an amino acid sequence at least about 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, 96%, 97%, 98%, 99% or more homologous to the amino acid sequence shown in SEQ ID NO: 3 and retains at least one functional activity of the polypeptide according to SEQ ID NO: 3.

Functional equivalents of a protein according to the invention can also be identified e.g. by screening combinatorial libraries of mutants, e.g. truncation mutants, of the protein of the invention for baking enzyme activity. In one embodiment, a variegated library of variants is generated by combinatorial mutagenesis at the nucleic acid level. A variegated library of variants can be produced by, for example, enzymatically ligating a mixture of synthetic oligonucleotides into gene sequences such that a degenerate set of potential protein sequences is expressible as individual polypeptides, or alternatively, as a set of larger fusion proteins (e.g., for phage display). There are a variety of methods that can be used to produce libraries of potential variants of the polypeptides of the invention from a degenerate oligonucleotide sequence. Methods for synthesizing degenerate oligonucleotides are known in the art (see, e.g., Narang (1983) Tetrahedron 39:3; Itakura et al. (1984) Annu. Rev. Biochem. 53:323; Itakura et al. (1984) Science 198:1056; Ike et al. (1983) Nucleic Acid Res. 11:477).

In addition, libraries of fragments of the coding sequence of a polypeptide of the invention can be used to generate a variegated population of polypeptides for screening a subsequent selection of variants. For example, a library of coding sequence fragments can be generated by treating a double stranded PCR fragment of the coding sequence of interest with a nuclease under conditions wherein nicking occurs only about once per molecule, denaturing the double stranded DNA, renaturing the DNA to form double stranded DNA which can include sense/antisense pairs from different nicked products, removing single stranded portions from reformed duplexes by treatment with S1 nuclease, and ligating the resulting fragment library into an expression vector. By this method, an expression library can be derived which encodes N-terminal and internal fragments of various sizes of the protein of interest.

Several techniques are known in the art for screening gene products of combinatorial libraries made by point mutations of truncation, and for screening cDNA libraries for gene products having a selected property. The most widely used

techniques, which are amenable to high through-put analysis, for screening large gene libraries typically include cloning the gene library into replicable expression vectors, transforming appropriate cells with the resulting library of vectors, and expressing the combinatorial genes under conditions in which detection of a desired activity facilitates
5 isolation of the vector encoding the gene whose product was detected. Recursive ensemble mutagenesis (REM), a technique which enhances the frequency of functional mutants in the libraries, can be used in combination with the screening assays to identify variants of a protein of the invention (Arkin and Yourvan (1992) Proc. Natl. Acad. Sci. USA 89:7811-7815; Delgrave et al. (1993) Protein Engineering
10 6(3):327-331).

In addition to the NBE 033 gene sequence shown in SEQ ID NO: 1, it will be apparent for the person skilled in the art that DNA sequence polymorphisms that may lead to changes in the amino acid sequence of the NBE 033 protein may exist within a given population. Such genetic polymorphisms may exist in cells from different
15 populations or within a population due to natural allelic variation. Allelic variants may also include functional equivalents.

Fragments of a polynucleotide according to the invention may also comprise polynucleotides not encoding functional polypeptides. Such polynucleotides may function as probes or primers for a PCR reaction.

20 Nucleic acids according to the invention irrespective of whether they encode functional or non-functional polypeptides, can be used as hybridization probes or polymerase chain reaction (PCR) primers. Uses of the nucleic acid molecules of the present invention that do not encode a polypeptide having a NBE 033 activity include, inter alia, (1) isolating the gene encoding the NBE 033 protein, or allelic variants
25 thereof from a cDNA library e.g. from other organisms than *A. niger*; (2) in situ hybridization (e.g. FISH) to metaphase chromosomal spreads to provide precise chromosomal location of the NBE 033 gene as described in Verma et al., Human Chromosomes: a Manual of Basic Techniques, Pergamon Press, New York (1988); (3) Northern blot analysis for detecting expression of NBE 033 mRNA in specific tissues
30 and/or cells and 4) probes and primers that can be used as a diagnostic tool to analyse the presence of a nucleic acid hybridisable to the NBE 033 probe in a given biological (e.g. tissue) sample.

Also encompassed by the invention is a method of obtaining a functional equivalent of a NBE 033 gene or cDNA. Such a method entails obtaining a
35 labelled probe that includes an isolated nucleic acid which encodes all or a portion of the sequence according to SEQ ID NO: 3 or a variant thereof; screening a nucleic acid

fragment library with the labelled probe under conditions that allow hybridisation of the probe to nucleic acid fragments in the library, thereby forming nucleic acid duplexes, and preparing a full-length gene sequence from the nucleic acid fragments in any labelled duplex to obtain a gene related to the NBE 033 gene.

5 In one embodiment, a NBE 033 nucleic acid of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to a nucleic acid sequence shown in SEQ ID NO: 1, SEQ ID NO: 2 or the complement thereof.

10 In another preferred embodiment a NBE 033 polypeptide of the invention is at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or more homologous to the amino acid sequence shown in SEQ ID NO: 3.

Host cells

15 In another embodiment, the invention features cells, e.g., transformed host cells or recombinant host cells that contain a nucleic acid encompassed by the invention. A "transformed cell" or "recombinant cell" is a cell into which (or into an ancestor of which) has been introduced, by means of recombinant DNA techniques, a
20 nucleic acid according to the invention. Both prokaryotic and eukaryotic cells are included, e.g., bacteria, fungi, yeast, and the like, especially preferred are cells from filamentous fungi, in particular *Aspergillus niger*.

A host cell can be chosen that modulates the expression of the inserted sequences, or modifies and processes the gene product in a specific, desired
25 fashion. Such modifications (e.g., glycosylation) and processing (e.g., cleavage) of protein products may facilitate optimal functioning of the protein.

Various host cells have characteristic and specific mechanisms for post-translational processing and modification of proteins and gene products.
Appropriate cell lines or host systems familiar to those of skill in the art of molecular
30 biology and/or microbiology can be chosen to ensure the desired and correct modification and processing of the foreign protein expressed. To this end, eukaryotic host cells that possess the cellular machinery for proper processing of the primary transcript, glycosylation, and phosphorylation of the gene product can be used. Such host cells are well known in the art.

35 Host cells also include, but are not limited to, mammalian cell lines such as CHO, VERO, BHK, HeLa, COS, MDCK, 293, 3T3, WI38, and choroid plexus

cell lines.

If desired, the polypeptides according to the invention can be produced by a stably-transfected cell line. A number of vectors suitable for stable transfection of mammalian cells are available to the public, methods for constructing such cell lines are also publicly known, e.g., in Ausubel et al. (supra).

Antibodies

The invention further features antibodies, such as monoclonal or polyclonal antibodies, that specifically bind NBE 033 proteins according to the invention.

As used herein, the term "antibody" (Ab) or "monoclonal antibody" (Mab) is meant to include intact molecules as well as antibody fragments (such as, for example, Fab and F(ab')₂ fragments) which are capable of specifically binding to NBE 033 protein. Fab and F(ab')₂ fragments lack the Fc fragment of intact antibody, clear more rapidly from the circulation, and may have less non-specific tissue binding of an intact antibody (Wahl et al., *J. Nucl. Med.* 24:316-325 (1983)). Thus, these fragments are preferred.

The antibodies of the present invention may be prepared by any of a variety of methods. For example, cells expressing the NBE 033 protein or an antigenic fragment thereof can be administered to an animal in order to induce the production of sera containing polyclonal antibodies. In a preferred method, a preparation of NBE 033 protein is prepared and purified to render it substantially free of natural contaminants. Such a preparation is then introduced into an animal in order to produce polyclonal antisera of greater specific activity.

In the most preferred method, the antibodies of the present invention are monoclonal antibodies (or NBE 033 protein binding fragments thereof). Such monoclonal antibodies can be prepared using hybridoma technology (Kohler et al., *Nature* 256:495 (1975); Kohler et al., *Eur. J. Immunol.* 6:511 (1976); Hammerling et al., In: *Monoclonal Antibodies and T-Cell Hybridomas*, Elsevier, N.Y., pp. 563-681 (1981)). In general, such procedures involve immunizing an animal (preferably a mouse) with a NBE 033 protein antigen or, with a NBE 033 protein expressing cell. The splenocytes of such mice are extracted and fused with a suitable myeloma cell line. Any suitable myeloma cell line may be employed in accordance with the present invention; however, it is preferably to employ the parent myeloma cell line (SP₂O), available from the American Type Culture Collection, Rockville, Maryland. After fusion, the resulting

hybridoma cells are selectively maintained in HAT medium, and then cloned by limiting dilution as described by Wands *et al.* (*Gastro-enterology* 80:225-232 (1981)). The hybridoma cells obtained through such a selection are then assayed to identify clones which secrete antibodies capable of binding the NBE 033 protein antigen. In general, the polypeptides can be coupled to a carrier protein, such as KLH, as described in Ausubel *et al.*, supra, mixed with an adjuvant, and injected into a host mammal.

In particular, various host animals can be immunized by injection of a polypeptide of interest. Examples of suitable host animals include rabbits, mice, guinea pigs, and rats. Various adjuvants can be used to increase the immunological response, depending on the host species, including but not limited to Freund's (complete and incomplete), adjuvant mineral gels such as aluminum hydroxide, surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, dinitrophenol, BCG (bacille Calmette-Guerin) and *Corynebacterium parvum*. Polyclonal antibodies are heterogeneous populations of antibody molecules derived from the sera of the immunized animals.

Such antibodies can be of any immunoglobulin class including IgG, IgM, IgE, IgA, IgD, and any subclass thereof. The hybridomas producing the mAbs of this invention can be cultivated *in vitro* or *in vivo*.

Once produced, polyclonal or monoclonal antibodies are tested for specific recognition of a NBE 033 polypeptide or functional equivalent thereof in an immunoassay, such as a Western blot or immunoprecipitation analysis using standard techniques, e.g., as described in Ausubel *et al.*, supra. Antibodies that specifically bind to NBE 033 proteins or functional equivalents thereof are useful in the invention. For example, such antibodies can be used in an immunoassay to detect NBE 033 in pathogenic or non-pathogenic strains of *Aspergillus* (e.g., in *Aspergillus* extracts).

Preferably, antibodies of the invention are produced using fragments of the NBE 033 polypeptides that appear likely to be antigenic, by criteria such as high frequency of charged residues. For example, such fragments may be generated by standard techniques of PCR, and then cloned into the pGEX expression vector (Ausubel *et al.*, supra). Fusion proteins may then be expressed in *E. coli* and purified using a glutathione agarose affinity matrix as described in Ausubel, *et al.*, supra. If desired, several (e.g., two or three) fusions can be generated for each protein, and each fusion can be injected into at least two rabbits. Antisera can be raised by injections in a series, typically including at least three booster injections. Typically, the antisera are checked for their ability to immunoprecipitate a recombinant NBE 033 polypeptide or functional equivalents thereof whereas unrelated proteins may serve as

a control for the specificity of the immune reaction.

Alternatively, techniques described for the production of single chain antibodies (U.S. Patent 4,946,778 and 4,704,692) can be adapted to produce single chain antibodies against a NBE 033 polypeptide or functional equivalents thereof. Kits
5 for generating and screening phage display libraries are commercially available e.g. from Pharmacia.

Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223, 409; PCT Publication No. WO 92/18619; PCT
10 Publication No. WO 91/17271; PCT Publication No. WO 20791; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs et al. (1991) *Bio/Technology* 9:1370-1372; Hay et al. (1992) *Hum. Antibod. Hybridomas* 3:81-85; Huse et al. (1989) *Science*
15 246:1275-1281; Griffiths et al. (1993) *EMBO J.* 12:725-734.

Polyclonal and monoclonal antibodies that specifically bind NBE 033 polypeptides of functional equivalents thereof can be used, for example, to detect expression of a NBE 033 gene or a functional equivalent thereof e.g. in another strain of *Aspergillus*. For example, NBE 033 polypeptide can be readily detected in
20 conventional immunoassays of *Aspergillus* cells or extracts. Examples of suitable assays include, without limitation, Western blotting, ELISAs, radioimmune assays, and the like.

By "specifically binds" is meant that an antibody recognizes and binds a particular antigen, e.g., a NBE 033 polypeptide, but does not substantially
25 recognize and bind other unrelated molecules in a sample.

Antibodies can be purified, for example, by affinity chromatography methods in which the polypeptide antigen is immobilized on a resin.

An antibody directed against a polypeptide of the invention (e.g., monoclonal antibody) can be used to isolate the polypeptide by standard techniques,
30 such as affinity chromatography or immunoprecipitation. Moreover, such an antibody can be used to detect the protein (e.g., in a cellular lysate or cell supernatant) in order to evaluate the abundance and pattern of expression of the polypeptide. The antibodies can also be used diagnostically to monitor protein levels in cells or tissue as part of a clinical testing procedure, e.g., to, for example, determine the efficacy of a given
35 treatment regimen or in the diagnosis of Aspergillosis..

Detection can be facilitated by coupling the antibody to a detectable

substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β -galactosidase, or acetylcholinesterase; examples of suitable
5 fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive materials include ^{125}I , ^{131}I , ^{35}S or ^3H .

10 Preferred epitopes encompassed by the antigenic peptide are regions that are located on the surface of the protein, e.g., hydrophilic regions. Hydrophobicity plots of the proteins of the invention can be used to identify hydrophilic regions.

The antigenic peptide of a protein of the invention comprises at least
15 7 (preferably 10, 15, 20, or 30) contiguous amino acid residues of the amino acid sequence of SEQ ID NO: 3 and encompasses an epitope of the protein such that an antibody raised against the peptide forms a specific immune complex with the protein.

Preferred epitopes encompassed by the antigenic peptide are regions of NBE 033 that are located on the surface of the protein, e.g., hydrophilic
20 regions, hydrophobic regions, alpha regions, beta regions, coil regions, turn regions and flexible regions.

Immunoassays

25 Qualitative or quantitative determination of a polypeptide according to the present invention in a biological sample can occur using any art-known method. Antibody-based techniques provide special advantages for assaying specific polypeptide levels in a biological sample.

In these, the specific recognition is provided by the primary antibody
30 (polyclonal or monoclonal) but the secondary detection system can utilize fluorescent, enzyme, or other conjugated secondary antibodies. As a result, an immunocomplex is obtained.

Accordingly, the invention provides a method for diagnosing whether a certain organism is infected with *Aspergillus* comprising the steps of:

- 35
- Isolating a biological sample from said organism suspected to be infected with *Aspergillus*,

- reacting said biological sample with an antibody according to the invention,
- determining whether immunocomplexes are formed.

5 Tissues can also be extracted, e.g., with urea and neutral detergent, for the liberation of protein for Western-blot or dot/slot assay. This technique can also be applied to body fluids.

Other antibody-based methods useful for detecting NBE 033 gene expression include immunoassays, such as the enzyme linked immunosorbent assay (ELISA) and the radioimmunoassay (RIA). For example, NBE 033-specific monoclonal antibodies can be used both as an immunoabsorbent and as an enzyme-labeled probe
10 to detect and quantify the NBE 033 protein. The amount of NBE 033 protein present in the sample can be calculated by reference to the amount present in a standard preparation using a linear regression computer algorithm. In another ELISA assay, two distinct specific monoclonal antibodies can be used to detect NBE 033 protein in a biological fluid. In this assay, one of the antibodies is used as the immuno-absorbent
15 and the other as the enzyme-labeled probe.

The above techniques may be conducted essentially as a "one-step" or "two-step" assay. The "one-step" assay involves contacting NBE 033 protein with immobilized antibody and, without washing, contacting the mixture with the labeled antibody. The "two-step" assay involves washing before contacting the mixture with the
20 labeled antibody. Other conventional methods may also be employed as suitable. It is usually desirable to immobilize one component of the assay system on a support, thereby allowing other components of the system to be brought into contact with the component and readily removed from the sample.

Suitable enzyme labels include, for example, those from the oxidase
25 group, which catalyze the production of hydrogen peroxide by reacting with substrate. Activity of an oxidase label may be assayed by measuring the concentration of hydrogen peroxide formed by the enzyme-labelled antibody/substrate reaction.

Besides enzymes, other suitable labels include radioisotopes, such as iodine (^{125}I , ^{127}I), carbon (^{14}C), sulphur (^{35}S), tritium (^3H), indium (^{112}In), and
30 technetium ($^{99\text{m}}\text{Tc}$), and fluorescent labels, such as fluorescein and rhodamine, and biotin.

Specific binding of a test compound to a NBE 033 polypeptide can be detected, for example, in vitro by reversibly or irreversibly immobilizing the NBE 033 polypeptide on a substrate, e.g., the surface of a well of a 96-well polystyrene microtitre
35 plate. Methods for immobilizing polypeptides and other small molecules are well known

in the art. For example, the microtitre plates can be coated with a NBE 033 polypeptide by adding the polypeptide in a solution (typically, at a concentration of 0.05 to 1 mg/ml in a volume of 1-100 ul) to each well, and incubating the plates at room temperature to 37 °C for 0.1 to 36 hours. Polypeptides that are not bound to the plate can be removed
5 by shaking the excess solution from the plate, and then washing the plate (once or repeatedly) with water or a buffer. Typically, the polypeptide is contained in water or a buffer. The plate is then washed with a buffer that lacks the bound polypeptide. To block the free protein-binding sites on the plates, the plates are blocked with a protein that is unrelated to the bound polypeptide. For example, 300 ul of bovine serum
10 albumin (BSA) at a concentration of 2 mg/ml in Tris-HCl is suitable. Suitable substrates include those substrates that contain a defined cross-linking chemistry (e.g., plastic substrates, such as polystyrene, styrene, or polypropylene substrates from Corning Costar Corp. (Cambridge, MA), for example) . If desired, a beaded particle, e.g., beaded agarose or beaded sepharose, can be used as the substrate.

15 Binding of the test compound to the polypeptides according to the invention can be detected by any of a variety of artknown methods. For example, a specific antibody can be used in an immunoassay. If desired, the antibody can be labeled (e.g., fluorescently or with a radioisotope) and detected directly (see, e.g., West and McMahon, J. Cell Biol. 74:264, 1977). Alternatively, a second antibody can be
20 used for detection (e.g., a labeled antibody that binds the Fc portion of an anti-AN97 antibody). In an alternative detection method, the NBE 033 polypeptide is labeled, and the label is detected (e.g., by labeling a NBE 033 polypeptide with a radioisotope, fluorophore, chromophore, or the like). In still another method, the NBE 033 polypeptide is produced as a fusion protein with a protein that can be detected
25 optically, e.g., green fluorescent protein (which can be detected under UV light). In an alternative method, the NBE 033 polypeptide can be covalently attached to or fused with an enzyme having a detectable enzymatic activity, such as horse radish peroxidase, alkaline phosphatase, α -galactosidase, or glucose oxidase. Genes encoding all of these enzymes have been cloned and are readily available for use by
30 those of skill in the art. If desired, the fusion protein can include an antigen, and such an antigen can be detected and measured with a polyclonal or monoclonal antibody using conventional methods. Suitable antigens include enzymes (e.g., horse radish peroxidase, alkaline phosphatase, and α -galactosidase) and non-enzymatic polypeptides (e.g., serum proteins, such as BSA and globulins, and milk proteins, such
35 as caseins).

Epitopes, antigens and immunogens.

In another aspect, the invention provides a peptide or polypeptide comprising an epitope-bearing portion of a polypeptide of the invention. The epitope of
5 this polypeptide portion is an immunogenic or antigenic epitope of a polypeptide of the invention. An "immunogenic epitope" is defined as a part of a protein that elicits an antibody response when the whole protein is the immunogen. These immunogenic epitopes are believed to be confined to a few loci on the molecule. On the other hand, a region of a protein molecule to which an antibody can bind is defined as an "antigenic
10 epitope." The number of immunogenic epitopes of a protein generally is less than the number of antigenic epitopes. See, for instance, Geysen, H. M. et al., Proc. Natl. Acad. Sci. USA 81:3998-4002 (1984).

As to the selection of peptides or polypeptides bearing an antigenic epitope (i.e., that contain a region of a protein molecule to which an antibody can bind),
15 it is well known in that art that relatively short synthetic peptides that mimic part of a protein sequence are routinely capable of eliciting an antiserum that reacts with the partially mimicked protein. See, for instance, Sutcliffe, J. G. et al., Science 219:660-666 (1984). Peptides capable of eliciting protein-reactive sera are frequently represented in the primary sequence of a protein, can be characterized by a set of simple chemical
20 rules, and are confined neither to immunodominant regions of intact proteins (i.e., immunogenic epitopes) nor to the amino or carboxyl terminals. Peptides that are extremely hydrophobic and those of six or fewer residues generally are ineffective at inducing antibodies that bind to the mimicked protein; longer, soluble peptides, especially those containing proline residues, usually are effective. Sutcliffe et al., supra,
25 at 661. For instance, 18 of 20 peptides designed according to these guidelines, containing 8-39 residues covering 75% of the sequence of the influenza virus hemagglutinin HAI polypeptide chain, induced antibodies that reacted with the HA1 protein or intact virus; and 12/12 peptides from the MuLV polymerase and 18/18 from the rabies glycoprotein induced antibodies that precipitated the respective proteins.

30 Antigenic epitope-bearing peptides and polypeptides of the invention are therefore useful to raise antibodies, including monoclonal antibodies, that bind specifically to a polypeptide of the invention. Thus, a high proportion of hybridomas obtained by fusion of spleen cells from donors immunized with an antigen epitope-bearing peptide generally secrete antibody reactive with the native protein. Sutcliffe et
35 al., supra, at 663. The antibodies raised by antigenic epitope bearing peptides or polypeptides are useful to detect the mimicked protein, and antibodies to different

peptides may be used for tracking the fate of various regions of a protein precursor which undergoes posttranslation processing. The peptides and anti-peptide antibodies may be used in a variety of qualitative or quantitative assays for the mimicked protein, for instance in competition assays since it has been shown that even short peptides (e.g., about 9 amino acids) can bind and displace the larger peptides in immunoprecipitation assays. See, for instance, Wilson, I.A. et al., Cell 37:767-778 at 777 (1984). The anti-peptide antibodies of the invention also are useful for purification of the mimicked protein, for instance, by adsorption chromatography using methods well known in the art.

Antigenic epitope-bearing peptides and polypeptides of the invention designed according to the above guidelines preferably contain a sequence of at least seven, more preferably at least nine and most preferably between about 15 to about 30 amino acids contained within the amino acid sequence of a polypeptide of the invention. However, peptides or polypeptides comprising a larger portion of an amino acid sequence of a polypeptide of the invention, containing about 30 to about 50 amino acids, or any length up to and including the entire amino acid sequence of a polypeptide of the invention, also are considered epitope-bearing peptides or polypeptides of the invention and also are useful for inducing antibodies that react with the mimicked protein. Preferably, the amino acid sequence of the epitope-bearing peptide is selected to provide substantial solubility in aqueous solvents (i.e., the sequence includes relatively hydrophilic residues and highly hydrophobic sequences are preferably avoided); and sequences containing proline residues are particularly preferred.

The epitope-bearing peptides and polypeptides of the invention may be produced by any conventional means for making peptides or polypeptides including recombinant means using nucleic acid molecules of the invention. For instance, a short epitope-bearing amino acid sequence may be fused to a larger polypeptide which acts as a carrier during recombinant production and purification, as well as during immunization to produce anti-peptide antibodies.

Epitope-bearing peptides also may be synthesized using known methods of chemical synthesis. For instance, Houghten has described a simple method for synthesis of large numbers of peptides, such as 10-20 mg of 248 different 13 residue peptides representing single amino acid variants of a segment of the HAI polypeptide which were prepared and characterized (by ELISA-type binding studies) in less than four weeks. Houghten, R. A., Proc. Natl. Acad. Sci. USA 82:5131-5135 (1985). This "Simultaneous Multiple Peptide Synthesis (SMPS)" process is further

described in U.S. Patent No. 4,631,211 to Houghten et al. (1986). In this procedure the individual resins for the solid-phase synthesis of various peptides are contained in separate solvent-permeable packets, enabling the optimal use of the many identical repetitive steps involved in solid-phase methods.

5 A completely manual procedure allows 500-1000 or more syntheses to be conducted simultaneously. Houghten et al., *supra*, at 5134.

Epitope-bearing peptides and polypeptides of the invention are used to induce antibodies according to methods well known in the art. See, for instance, Sutcliffe et al., *supra*; Wilson et al., *supra*; Chow, M. et al., *Proc. Natl. Acad. Sci. USA* 10 82:910-914; and Bittle, F.J. et al., *J. Gen. Virol.* 66:2347-2354 (1985).

Generally, animals may be immunized with free peptide; however, anti-peptide antibody titer may be boosted by coupling of the peptide to a macromolecular carrier, such as keyhole limpet hemocyanin (KLH) or tetanus toxoid. For instance, peptides containing cysteine may be coupled to carrier using a linker 15 such as maleimidobenzoyl-N-hydroxysuccinimide ester (MBS), while other peptides may be coupled to carrier using a more general linking agent such as glutaraldehyde.

Animals such as rabbits, rats and mice are immunized with either free or carriercoupled peptides, for instance, by intraperitoneal and/or intradermal injection of emulsions containing about 100 ug peptide or carrier protein and Freund's adjuvant.

20 Several booster injections may be needed, for instance, at intervals of about two weeks, to provide a useful titer of anti-peptide antibody which can be detected, for example, by ELISA assay using free peptide adsorbed to a solid surface. The titer of anti-peptide antibodies in serum from an immunized animal may be increased by selection of anti-peptide antibodies, for instance, by adsorption to the peptide on a solid 25 support and elution of the selected antibodies according to methods well known in the art.

Immunogenic epitope-bearing peptides of the invention, i.e., those parts of a protein that elicit an antibody response when the whole protein is the immunogen, are identified according to methods known in the art. For instance, 30 Geysen et al., 1984, *supra*, discloses a procedure for rapid concurrent synthesis on solid supports of hundreds of peptides of sufficient purity to react in an enzyme-linked immunosorbent assay. Interaction of synthesized peptides with antibodies is then easily detected without removing them from the support. In this manner a peptide bearing an immunogenic epitope of a desired protein may be identified routinely by one 35 of ordinary skill in the art. For instance, the immunologically important epitope in the coat protein of foot-and-mouth disease virus was located by Geysen et al. with a

resolution of seven amino acids by synthesis of an overlapping set of all 208 possible hexapeptides covering the entire 213 amino acid sequence of the protein. Then, a complete replacement set of peptides in which all 20 amino acids were substituted in turn at every position within the epitope were synthesized, and the particular amino acids conferring specificity for the reaction with antibody were determined. Thus, peptide analogs of the epitope-bearing peptides of the invention can be made routinely by this method. U.S. Patent No. 4,708,781 to Geysen (1987) further describes this method of identifying a peptide bearing an immunogenic epitope of a desired protein.

Further still, U.S. Patent No. 5,194,392 to Geysen (1990) describes a general method of detecting or determining the sequence of monomers (amino acids or other compounds) which is a topological equivalent of the epitope (i.e., a "mimotope") which is complementary to a particular paratope (antigen binding site) of an antibody of interest. More generally, U.S. Patent No. 4,433,092 to Geysen (1989) describes a method of detecting or determining a sequence of monomers which is a topographical equivalent of a ligand which is complementary to the ligand binding site of a particular receptor of interest. Similarly, U.S. Patent No. 5,480,971 to Houghten, R. A. et al. (1996) on Peralkylated Oligopeptide Mixtures discloses linear C1-C7-alkyl peralkylated oligopeptides and sets and libraries of such peptides, as well as methods for using such oligopeptide sets and libraries for determining the sequence of a peralkylated oligopeptide that preferentially binds to an acceptor molecule of interest. Thus, non-peptide analogs of the epitope-bearing peptides of the invention also can be made routinely by these methods.

Use of NBE 033 baking enzymes in industrial processes

The invention also relates to the use of the NBE 033 baking enzyme according to the invention in a selected number of industrial processes. Despite the long term experience obtained with these processes, the baking enzyme according to the invention features a number of significant advantages over the enzymes currently used. Depending on the specific application, these advantages can include aspects like lower production costs, higher specificity towards the substrate, less antigenic, less undesirable side activities, higher yields when produced in a suitable microorganism, more suitable pH and temperature ranges, better tastes of the final product as well as food grade and kosher aspects.

CLAIMS

1. An isolated polynucleotide hybridisable to a polynucleotide according to SEQ ID NO:1 or SEQ ID NO: 2.
- 5 2. An isolated polynucleotide according to claim 1 hybridisable under high stringency conditions to a polynucleotide according to SEQ ID NO:1 or SEQ ID NO: 2.
3. An isolated polynucleotide according to claims 1 or 2 obtainable from a filamentous fungus.
- 10 4. An isolated polynucleotide according to claim 3 obtainable from *A. niger*.
5. An isolated polynucleotide encoding a polypeptide comprising an amino acid sequence according to SEQ ID NO: 3 or functional equivalents thereof.
6. An isolated polynucleotide encoding at least one functional domain of a polypeptide according to SEQ ID NO: 3 or functional equivalents thereof.
- 15 7. An isolated polynucleotide comprising a nucleotide sequence according to SEQ ID NO: 1 or SEQ ID NO: 2 or functional equivalents thereof.
8. An isolated polynucleotide according to SEQ ID NO: 1 or SEQ ID NO: 2.
9. A vector comprising a polynucleotide sequence according to claims 1 to 8.
10. A vector according to claim 9 wherein said polynucleotide sequence according to claims 1 to 8 is operatively linked with regulatory sequences suitable for expression of said polynucleotide sequence in a suitable host cell.
- 20 11. A vector according to claim 10 wherein said suitable host cell is a filamentous fungus.
12. A method for manufacturing a polynucleotide according to claims 1 – 8 or a vector according to claims 9 to 11 comprising the steps of culturing a host cell transformed with said polynucleotide or said vector and isolating said polynucleotide or said vector from said host cell.
- 25 13. An isolated polypeptide according to SEQ ID NO: 3 or functional equivalents

thereof.

14. An isolated polypeptide according to claim 13 obtainable from *Aspergillus niger*.
- 5 15. An isolated polypeptide obtainable by expressing a polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11 in an appropriate host cell, e.g. *Aspergillus niger*.
16. Recombinant baking enzyme comprising a functional domain of a NBE 033 polypeptide.
- 10 17. A method for manufacturing a polypeptide according to claims 13 to 16 comprising the steps of transforming a suitable host cell with an isolated polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11, culturing said cell under conditions allowing expression of said polynucleotide and optionally purifying the encoded polypeptide from said cell or culture medium.
- 15 18. A recombinant host cell comprising a polynucleotide according to claims 1 to 8 or a vector according to claims 9 to 11.
19. A recombinant host cell expressing a polypeptide according to claims 13 to 16.
20. Purified antibodies reactive with a polypeptide according to claims 13 to 16.
21. Fusion protein comprising a polypeptide sequence according to claims 13 to 16.

ABSTRACT

The invention relates to a newly identified polynucleotide sequence comprising a gene that encodes a novel baking enzyme. The invention features the full length nucleotide sequence of the novel gene, the cDNA sequence comprising the full
5 length coding sequence of the novel baking enzyme as well as the amino acid sequence of the full-length functional protein and functional equivalents thereof. The invention also relates to methods of using these enzymes in industrial processes and methods of diagnosing fungal infections. Also included in the invention are cells transformed with a polynucleotide according to the invention and cells wherein a baking
10 enzyme according to the invention is genetically modified to enhance its activity and/or level of expression.

SEQUENCE LISTING

<110> DSM N.V.

<120> Novel baking enzyme, NBE033, and uses thereof

<130> 21083EP/P0/

<160> 3

<170> XML to WIPO ST.25 Converter - <http://www.biomax.de/>

<210> 1

<211> 1881

<212> DNA

<213> *Aspergillus niger*

<400> 1

ccgagtgate	cgcgacttct	gggtcagctt	tggtctgcgg	gtccaccag	tgccgtcatc	60
gcacgcgaga	cccgccccgg	atgcttagct	gaagccggca	tagtccatcc	ccgctcgggc	120
gtgatgcccc	acgggtcactg	gaggagaggt	gcaggaggga	tgccgttgca	tgaatcaagc	180
ccgggggtttg	actttcatcc	gctcgtttcc	tactggcggg	ttcgccctct	ccatcgaagc	240
cacggatcct	tccatccgga	tcctggcaga	acagtgagga	agcagagctt	ggttatagta	300
gaaattatta	ataccgagct	ggtctgcccc	tttttcccaa	acottccctc	tttccatccc	360
tctcgccctcg	cacccccctt	atcctccctc	ccgccatgta	tatccctctg	gtgctgcttc	420
tgggccgcgag	cctgttccat	ggcgcaacgg	cgctgcccac	gcccggctcc	acgcccaccc	480
cgcccagcca	ggatccctgg	tacagtgcgc	ccgaggggctt	cgaggaggct	gatcccgggtg	540
ccatcctgcg	cgtgoggccc	gcgcccggca	acttgaccgt	ggtagtgggc	aatgcgtcgg	600
cggcctacaa	catcctctac	cgactacag	acagtacgta	caagccctcc	tgggctgtga	660
ccaccctgct	ggtgcccccc	gtggccgctt	ccgcccgcgt	caaccagagt	gtcctgctct	720
cccaccagat	cgctacagat	tcgttcgacg	tcaatgccag	tcccagctac	gccatgtaca	780
ccagcccgcc	ctccgatatt	atcctcgccc	tgcaagcgcg	ctgggtcggt	aacgtccccg	840
attacgaggg	ccccaatgcc	tctttcacgg	ccggtgtgca	gtccggccat	gccaccctcg	900
actcgggtccg	cagcgtgctc	gcctccggat	tcggcctgaa	cgaggacgcc	cagtacgctc	960
tggtgggttta	ctctggcggt	gccttgggca	gogaatgggc	tgctgaactg	cagatgcaat	1020
acgtcccga	gttgaacatt	gccggtcttg	ccgtgggttg	tctcactccc	aatgttacca	1080
gcgtcatgga	cacggtgacc	tcgaccatca	gtgcgggact	catccccgcc	gccgccctgg	1140
gtctgtcgag	ccagcacccc	gagacctacg	agttcatcct	cagccagctc	aagacgacgg	1200
gaccctacaa	ccgcacagga	ttcctagccg	ccaaggacct	gaccctgtcc	gaggcgagg	1260
tcttctacgc	cttccagaac	atcttcgatt	actttgtcaa	cggatcgggc	acgttccagg	1320
cggaggtggt	gcagaaggcg	ctgaaccagg	acggatacat	gggtaccat	gggttcccgc	1380
agatgccggt	gctcgcgtac	aaggctattc	acgatgagat	cagtcccatc	caggatacgg	1440
atcgcgtgat	caagcgctac	tgtggtcttg	gattgaacat	cttgtatgag	cggaacacca	1500
tcggtggcca	ctcggcagag	caggtgaatg	gcaacgccag	ggcgtggaac	tggttgacga	1560
gcattttcga	cggaacgtat	gcgcagcagt	acaagaccga	ggggtgcacg	atccgcaatg	1620
tcactctgaa	cacgacttcc	tccgtttatt	agagaggggg	ctgttggtat	gtgaataatg	1680
ctgaagatgg	ctgtgtatgg	acggtccgct	ctcctgtata	gtaatgggct	aatgcatgcg	1740
gttccatgaa	catggtacga	aagattagat	tatgtatata	gtgtggaagt	ggtaatgatg	1800
atataatata	tgtctatgat	ctatgttcc	gctattctat	acaaacgcga	ccattcacag	1860
aatgactact	ggcacatctg	c				1881

<210> 2

<211> 1257

<212> DNA

<213> *Aspergillus niger*

<220>
 <221> CDS
 <222> (1)..(1257)

```

<400> 2
atg tat atc ccc tcg gtg ctg ctt ctg gcc gcg agc ctg ttc cat ggc      48
Met Tyr Ile Pro Ser Val Leu Leu Leu Ala Ala Ser Leu Phe His Gly
1          5          10          15
gca acg gcg ctg ccc acg ccc ggc tcc acg ccc atc ccg ccc agc cag      96
Ala Thr Ala Leu Pro Thr Pro Gly Ser Thr Pro Ile Pro Pro Ser Gln
          20          25          30
gat ccc tgg tac agt gcg ccc gag ggc ttc gag gag gct gat ccc ggt      144
Asp Pro Trp Tyr Ser Ala Pro Glu Gly Phe Glu Glu Ala Asp Pro Gly
          35          40          45
gcc atc ctg cgc gtg cgg ccc gcg ccc ggc aac ttg acc gtg gta gtg      192
Ala Ile Leu Arg Val Arg Pro Ala Pro Gly Asn Leu Thr Val Val Val
          50          55          60
ggc aat gcg tcg gcg gcc tac aac atc ctc tac cgc act aca gac agt      240
Gly Asn Ala Ser Ala Ala Tyr Asn Ile Leu Tyr Arg Thr Thr Asp Ser
65          70          75          80
cag tac aag ccc tcc tgg gct gtg acc acc ctg ctg gtg ccc ccc gtg      288
Gln Tyr Lys Pro Ser Trp Ala Val Thr Thr Leu Leu Val Pro Pro Val
          85          90          95
gcc gcc tcc gcc gcc gtc aac cag agt gtc ctg ctc tcc cac cag atc      336
Ala Ala Ser Ala Ala Val Asn Gln Ser Val Leu Leu Ser His Gln Ile
          100          105          110
gcc tac gat tcg ttc gac gtc aat gcc agt ccc agc tac gcc atg tac      384
Ala Tyr Asp Ser Phe Asp Val Asn Ala Ser Pro Ser Tyr Ala Met Tyr
          115          120          125
acc agc ccg ccc tcc gat att atc ctc gcc ctg cag cgc ggc tgg ttc      432
Thr Ser Pro Pro Ser Asp Ile Ile Leu Ala Leu Gln Arg Gly Trp Phe
          130          135          140
gtt aac gtc ccc gat tac gag ggc ccc aat gcc tct ttc acc gcc ggt      480
Val Asn Val Pro Asp Tyr Glu Gly Pro Asn Ala Ser Phe Thr Ala Gly
          145          150          155          160
gtg cag tcc ggc cat gcc acc ctc gac tcg gtc cgc agc gtg ctc gcc      528
Val Gln Ser Gly His Ala Thr Leu Asp Ser Val Arg Ser Val Leu Ala
          165          170          175
tcc gga ttc ggc ctg aac gag gac gcc cag tac gct ctg tgg ggt tac      576
Ser Gly Phe Gly Leu Asn Glu Asp Ala Gln Tyr Ala Leu Trp Gly Tyr
          180          185          190
tct ggc ggt gcc ttg gcc agc gaa tgg gct gct gaa ctg cag atg caa      624
Ser Gly Gly Ala Leu Ala Ser Glu Trp Ala Ala Glu Leu Gln Met Gln
          195          200          205
tac gct ccc gag ttg aac att gcc ggt ctg gcc gtg ggt ggt ctc act      672
Tyr Ala Pro Glu Leu Asn Ile Ala Gly Leu Ala Val Gly Gly Leu Thr
          210          215          220
ccc aat gtt acc agc gtc atg gac acg gtg acc tcg acc atc agt gcg      720
Pro Asn Val Thr Ser Val Met Asp Thr Val Thr Ser Thr Ile Ser Ala
          225          230          235          240
gga ctc atc ccc gcc gcc gcc ctg ggt ctg tcg agc cag cac ccc gag      768
Gly Leu Ile Pro Ala Ala Ala Leu Gly Leu Ser Ser Gln His Pro Glu
          245          250          255
acc tac gag ttc atc ctc agc cag ctc aag acg acg gga ccc tac aac      816
Thr Tyr Glu Phe Ile Leu Ser Gln Leu Lys Thr Thr Gly Pro Tyr Asn
          260          265          270

```

cgc aca gga ttc cta gcc gcc aag gac ctg acc ctg tcc gag gcg gag	864
Arg Thr Gly Phe Leu Ala Ala Lys Asp Leu Thr Leu Ser Glu Ala Glu	
275 280 285	
gtc ttc tac gcc ttc cag aac atc ttc gat tac ttt gtc aac gga tcg	912
Val Phe Tyr Ala Phe Gln Asn Ile Phe Asp Tyr Phe Val Asn Gly Ser	
290 295 300	
gcc acg ttc cag gcg gag gtg gtg cag aag gcg ctg aac cag gac gga	960
Ala Thr Phe Gln Ala Glu Val Val Gln Lys Ala Leu Asn Gln Asp Gly	
305 310 315 320	
tac atg ggc tac cat ggg ttc ccg cag atg ccg gtg ctc gcg tac aag	1008
Tyr Met Gly Tyr His Gly Phe Pro Gln Met Pro Val Leu Ala Tyr Lys	
325 330 335	
gct att cac gat gag atc agt ccc atc cag gat acg gat cgc gtg atc	1056
Ala Ile His Asp Glu Ile Ser Pro Ile Gln Asp Thr Asp Arg Val Ile	
340 345 350	
aag cgc tac tgt ggt ctg gga ttg aac atc ttg tat gag cgg aac acc	1104
Lys Arg Tyr Cys Gly Leu Gly Leu Asn Ile Leu Tyr Glu Arg Asn Thr	
355 360 365	
atc ggt ggc cac tcg gca gag cag gtg aat ggc aac gcc agg gcg tgg	1152
Ile Gly Gly His Ser Ala Glu Gln Val Asn Gly Asn Ala Arg Ala Trp	
370 375 380	
aac tgg ttg acg agc att ttc gac gga acg tat gcg cag cag tac aag	1200
Asn Trp Leu Thr Ser Ile Phe Asp Gly Thr Tyr Ala Gln Gln Tyr Lys	
385 390 395 400	
acc gag ggg tgc acg atc cgc aat gtc act ctg aac acg act tcc tcc	1248
Thr Glu Gly Cys Thr Ile Arg Asn Val Thr Leu Asn Thr Thr Ser Ser	
405 410 415	
gtt tat tag	
Val Tyr	1257

<210> 3
 <211> 418
 <212> PRT
 <213> *Aspergillus niger*

<400> 3

Met Tyr Ile Pro Ser Val Leu Leu Leu Ala Ala Ser Leu Phe His Gly	
1 5 10 15	
Ala Thr Ala Leu Pro Thr Pro Gly Ser Thr Pro Ile Pro Pro Ser Gln	
20 25 30	
Asp Pro Trp Tyr Ser Ala Pro Glu Gly Phe Glu Glu Ala Asp Pro Gly	
35 40 45	
Ala Ile Leu Arg Val Arg Pro Ala Pro Gly Asn Leu Thr Val Val Val	
50 55 60	
Gly Asn Ala Ser Ala Ala Tyr Asn Ile Leu Tyr Arg Thr Thr Asp Ser	
65 70 75 80	
Gln Tyr Lys Pro Ser Trp Ala Val Thr Thr Leu Leu Val Pro Pro Val	
85 90 95	
Ala Ala Ser Ala Ala Val Asn Gln Ser Val Leu Leu Ser His Gln Ile	
100 105 110	
Ala Tyr Asp Ser Phe Asp Val Asn Ala Ser Pro Ser Tyr Ala Met Tyr	
115 120 125	
Thr Ser Pro Pro Ser Asp Ile Ile Leu Ala Leu Gln Arg Gly Trp Phe	
130 135 140	
Val Asn Val Pro Asp Tyr Glu Gly Pro Asn Ala Ser Phe Thr Ala Gly	

[illegible]